

**STATISTIČKI PRAKTIKUM**

## DRUGI KOLOKVIJ

**Napomena:**

Prije početka rješavanja kolokvija instalirajte i učitajte pakete BSDA i HSAUR2, npr.

```
install.packages("HSAUR2")
```

```
library("HSAUR2")
```

Rješenja trebate kopirati u Word dokument. Komentare možete pisati u isti dokument ili na papir. Po završetku pisanja, trebate poslati dokument na

```
statistika@mathos.hr
```

i u predmetu poruke napisati svoje ime i prezime.

**ZADATAK 1:** [2+3+3=8 boda]

Odgovorite na sljedeća pitanja:

- Što je rang podatka  $x_i$  u nizu  $(x_1, \dots, x_n)$  u kojem su svi elementi međusobno različiti? Odredite rangove podataka u nizu (2.5, 1.1, 0.2, 4.5).
- Definirajte koeficijent korelacije slučajnih varijabli  $X$  i  $Y$ . Objasnite značenje vrijednosti 1,  $-1$  i 0 koeficijenta korelacije.
- Objasnite kako možemo testirati hipotezu o jednakosti proporcija u dva zavisna slučajna uzorka.

**ZADATAK 2:** [4+4+4+4=16 bodova]

Baza podataka `USmelanoma` iz paketa `HSAUR2` sadrži podatke o stopi smrtnosti od zloćudnog melanoma na 10 milijuna stanovnika za 49 američkih država. Varijabla `ocean` označava ima li država izlaz na ocean ili ne.

- Potvrdite da možemo pretpostaviti normalnu distribuiranost za stopu smrtnosti oceanskih država (`ocean=="yes"`) i neoceanskih (`ocean=="no"`).
- Možemo li tvrditi da je stopa smrtnosti veća za oceanske države na razini značajnosti 0.05?
- Je li stopa smrtnosti varijabilnija u oceanskim državama ( $\alpha = 0.05$ )?
- Postoji li razlika u distribucijama stope smrtnosti u dvije promatrane kategorije ( $\alpha = 0.05$ ).

**ZADATAK 3:** [5 bodova]

Šef tvrtke želi ukinuti pauzu za kavu svojim radnicima jer smatra da to neće smanjiti njihovu produktivnost. U bazi podataka `Coffee` iz paketa `BSDA` nalaze se podaci o produktivnosti radnika. Svakom radniku produktivnost je mjerena jednog dana kada je koristio pauzu za kavu (`With`) i jednog dana kada je nije koristio (`Without`). Pod pretpostavkom normalne distribuiranosti obilježja, možemo li na razini značajnosti 0.05 tvrditi da nedostatak pauze smanjuje produktivnost?

**ZADATAK 4:** [4+2=6 bodova]

Jane Austen je preminula prije nego je završila svoj posljednji roman. Imitator je pokušao dovršiti njezin roman. Sljedećom tablicom prikazane su frekvencije nekoliko čestih riječi u engleskom jeziku u slučajno odabranom dijelu teksta koji je napisala Austen, odnosno imitator.

	a	an	this	that	with	without
Austen	434	62	86	236	161	38
Imitator	83	29	15	22	43	4

Testirajte jesu li učestalosti odabranih riječi jednake kod Austen i kod imitatora, tj. je li imitator uspio dobro iskopirati stil Jane Austen ( $\alpha = 0.05$ ). Odredite empirijske distribucije promatranih riječi posebno za Austen i imitatora.

**ZADATAK 5:** [5 bodova]

Miš i hrčak testirani su u laboratorijskom labirintu. Miš je uspješno pronašao izlaz 12 puta od 100 pokušaja, dok je hrčak uspio u 10 puta od 70 pokušaja. Možemo li tvrditi da se hrčak bolje snalazi u labirintu ( $\alpha = 0.05$ )?

**ZADATAK 6:** [4+4+2=10 bodova]

Baza podataka `Verbal` iz paketa `BSDA` sadrži rezultate testova usmenog izražavanja (varijabla `verbal`) i broj posuđenih knjiga u knjižnici (varijabla `number`) za 15 osnovnoškolaca.

- Možemo li tvrditi da više posuđenih knjiga ukazuje na veći rezultat na testu ( $\alpha = 0.05$ )?
- Procijenite parametre linearnog modela koji opisuje rezultat testa u ovisnosti o broju posuđenih knjiga. Kako ocjenjujete jakost modela?
- Pod pretpostavkom da je model valjan, za koliko će se povećati rezultat testa sa svakom posuđenom knjigom?